PH US 010451 WO MAT. DOSSIER

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 August 2001 (09.08.2001)

PCT

(10) International Publication Number
WO 01/58145 A2

(51) International Patent Classification[7]: H04N 5/00

(21) International Application Number: PCT/EP01/00816

(22) International Filing Date: 25 January 2001 (25.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/498,271 4 February 2000 (04.02.2000) US

(71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor: SCHAFFER, James, D.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: GROENENDAAL, Antonius, W., M.; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States *(national)*: JP, KR.

(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/58145 A2

(54) Title: ADAPTIVE TV PROGRAM RECOMMENDER

(57) Abstract: A system for recommending television programs makes use of probabilistic calculations and a viewer profile to create a recommendation. The probabilistic calculations preferably are in the form of Bayesian classifier theory. Modifications to classical Bayesian classifier theory are proposed.

Adaptive tv program recommender

I. BACKGROUND OF THE INVENTION

A. Field of the Invention

The invention relates to recommending television shows based on a user profile.

B. Related Art

'U.S. Pat. No. 5,758,259 shows a method for identifying a preferred television program based on a "correlation" between the program and predetermined characteristics of a user profile. The term "correlation" as used in the patent does not appear to relate to the mathematical concept of correlation, but rather is a very simple algorithm for assessing some similarity between a profile and a program.

II. SUMMARY OF THE INVENTION

It is an object of the invention to improve techniques of automatic program recommendation.

This object is achieved by using a probabilistic calculation, based on a viewer profile. The probabilistic calculation is preferably based on Bayesian classifier theory.

The object is further achieved by maintaining a local record of a viewer history. The local record is preferably incrementally updatable. The local record is advantageous for privacy reasons, and can be contrasted with methods such as collaborative filtering, which would require viewer history information to be uploaded to a central location. The use of incremental updates is advantageous in minimizing storage requirements.

It is a still further object of the invention to improve the classical Bayesian classifier technique.

In one embodiment, this object is achieved by noise filtering.

In another embodiment, this object is achieved by applying a modified Bayesian classifier technique to non-independent feature values.

Further objects and advantages of the invention will be described in the following.

Bayesian classifiers are discussed in general in the text book Duda & Hart, Pattern Recognition and Scene Analysis (John Wiley & Sons 1973). An application of Bayesian classifiers to document retrieval is discussed in D. Billsus & M. Pazzani, Learning Probabilistic User Models", *http://www.dkfi.uin-sb.de/~bauer/um-ws/Final-Versions/Billsus/ProbUserModels.html*

III. BRIEF DESCRIPTION OF THE DRAWING

The invention will now be described by way of non-limiting example with reference to the following drawings.

Fig. 0 shows a system on which the invention may be used.

Fig. 1 shows major elements of an adaptive recommender.

Fig. 2 shows pseudo code for a viewing history generator.

Fig. 3 shows a table of key fields.

Fig. 4 shows a viewer profile.

Fig. 5a shows a prior probability calculation.

Fig. 5b shows a conditional probability calculation.

Fig. 5c shows a posterior probability calculation.

IV. DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

Fig 0 illustrates hardware for implementing the invention. The hardware will typically have a display 1; some type of processor 2; some type of user entry device 4 connected to the processor via some type of connection 3; and some type of link 5 for receiving data, such as television programming or Electronic Programming Guide ("EPG") data. The display 1 will commonly be a television screen, but could be any other type of display device. The processor 2 may be a set top box, a PC, or any other type of data processing device, so long as it has sufficient processing power. The user entry device 4 may be a remote and the connection 3 may be an infrared connection. If the processor is a PC, the user entry device will commonly be at least plural, e.g. a keyboard and mouse. The user

entry device may also be touch sensitivity on the display. The connection 5 to the outside world could be an antenna, cable, a phone line to the internet, a network connection, or any other data link. Equally well, connection 5 could connect to a memory device or several memory devices.

Fig. 1 illustrates major elements of an embodiment of an adaptive recommender. These elements preferably reside as software and data in a medium 110 readable by a data processing device such as CPU 2. The elements include a viewing history data structure 101 that gives input to profiler software 102. The profiler software in turn produces the viewer profile 103. The terms "user profile" and "viewer profile" shall be used interchangeably herein. The viewer profile serves as input to recommender software 104. The recommender software also uses, as input, the EPG data structure 105, that contains features describing each show such as title, channel, start time and the like. An output of the recommender 104 appears on a user interface 106 where a user can interact with it.

This viewer history data structure includes selected records from the EPG database. The EPG databases are commercially available, for instance from Tribune Media Services Those of ordinary skill in the art may devise other formats, possibly with finer shades of description. The selected records minimally correspond to TV shows watched by the viewer. It is assumed that these records have been deposited in the viewing history by software that is part of the user interface and knows what shows the viewer has viewed. Preferably, the software would allow recording of a user watching more than one show in a given time interval, as users often do switch back and forth during commercials and so forth. Preferably, also, the software records a program as watched regardless of how long it was watched; and if a show was watched or whether it was taped for later viewing.

The preferred viewing history format assumes the presence of both positive and negative records in the viewing history. This is needed because the goal is to learn to differentiate between the features of shows that are liked and those not liked. Fig. 2 shows pseudo code for collecting the viewing history.

Let the notation $C+$ denote the set of positive (i.e., watched) shows and $C-$ denote the negative (i.e., not watched) shows.

The viewer profile includes a number of feature value counts. These counts will be incremented whenever new entries are deposited in the viewer history. Usually, each program will have several feature values. Accordingly, the deposit of a program in the viewer history will cause the update of counts associated with all feature values associated with that program.

BNSDOCID: <WO 0158145A2 >

The incremental updatability of this type of history is advantageous because it allows for ongoing adaptation of the viewer history without a large amount of storage or computing effort being required.

In addition to the count of the number of positive and negative entries (k(C+), k(C-)), a count of occurrences of individual features will also be kept among the positive and negative examples (k(fi|C+), k(fi|C-)) where fi denotes feature i and k(fi|C+) denotes the number of shows in set C+ that possess feature fi. The feature set will include entries in the EPG records extracted from selected key fields, an example of which is shown in Table 1 which is Fig. 3.

A partial example of an embodiment of such counts is presented in table 2, shown in Fig. 4 to illustrate the idea. The list is presented in the Figure in six columns to save space, but in fact the list has only three columns, with the later part of the table being presented next to the earlier part. Each row of the column has four pieces of data: a feature type and a feature value in the first column, a positive count in the second column, and a negative count in the third column. The positive count indicates the number of times a program having that feature value has been watched. The negative count indicates the number of times a program having that feature value has not been watched.

A television program schedule normally includes several, if not many, programs for every time slot in every day. Normally, the user will only watch one or two of the programs in any given slot. If the viewer profile contains a list of ALL the programs not watched, the number of programs not watched will far exceed the number of programs watched. It may be desirable to create a method for sampling the programs not watched. For instance, as the processor assembles the viewer profile, the processor may chose a single not watched program at random from the weekly schedule as a companion for each watched program, as suggested in the pseudocode of Fig. 2. This design will attempt to keep the number of positive and negative entries in the viewing history about equal so as not to unbalance the Bayesian prior probability estimates, discussed below.

It is not generally desirable to choose a companion program from the same time slot as a watched program. Experiment has shown that the combined time and day feature value is typically the strongest or one of the strongest predictors of whether a particular program will be preferred. Thus another program at the same time as the watched program may well be a second or third choice program, while a program at a totally different time may be very undesirable. Accordingly, it is preferred to choose the companion program at random from the program schedule of the entire week that includes the watched program.

PNSDOCID: WO 0158145A2 I

Since time and day feature values for a program are often so important in determining whether a program will be of interest to a user, it is typically undesirable to consider two programs of identical content to be the same if they are shown on different days and/or at different times. In other words a particular episode of a series may be strongly preferred if it is shown at 8 p.m. on Tuesday, while the same episode of the same series may be completely undesirable if it is shown at 10 a.m. on Monday. Thus the episode at 10 a.m. should be considered a different program from the episode at 8 p.m., even though the content of the two are identical.

As more and more shows are viewed, the length of the profile will tend to grow larger and larger. To combat this, and to keep the focus on features that are effective discriminators, the following are recommended:

periodic reviews of the features in the viewer profile, and

removal of words that appear to be frequent and not very discriminating.

In general, those of simple tastes, e.g. those who only like to watch football, will be fairly easy to recommend for after taking of a viewer profile for a relatively short time. For those of more complex preferences, it will take longer for the viewer history to be sufficiently meaningful to make good recommendations. These latter people, however, are those who are probably most in need of a recommendation.

In the final analysis, viewer histories will always be ambiguous. Recommendations of shows based on such histories will always contain a margin for error. The recommendations can at best be said to have some probability of being correct. Therefore probabilistic calculations are useful in analyzing viewer profile data to make recommendations.

The preferred embodiment of the recommender uses a simple Bayesian classifier using prior and conditional probability estimates derived from the viewer profile. How recommendations are shown to viewers is not defined here, yet it will be assumed that one can capture the viewer's response to them, at least observing whether or not they were watched.

Below will be discussed a 2-class Bayesian decision model. The two classes of TV shows of interest are:

C1 – shows that interest the viewer

C2 – shows that do not interest the viewer

Other classes might be used showing more shades of interest or lack thereof.

In contrast with the classes of interest listed above, viewing history obtains information only on the classes:

C+ - shows the viewer watched

C- - shows the viewer did not watch

Determining which shows a user watched or did not watch is outside of the scope of this application. The user might enter a manual log of which shows s/he watched. Alternatively, hardware might record the user's watching behavior. Those of ordinary skill in the art might devise numerous techniques for this. It should be possible to consider shows as watched even if they are watched only for a short time, as a user may be switching back and forth between several shows, trying to keep track of all of them.

Inferences may be made about classes *C1* and *C2* based on observations, but these inferences will always contain an element of uncertainty. The Bayesian model will compute the prior probabilities *P(C+)* and *P(C-)* directly from the counts in the viewer profile in accordance with Fig. 5a. In other words, the assumption will be that shows not watched are those the viewer is not interested in, and that the shows watched are the ones that the viewer is interested in.

The conditional probabilities, that a given feature, *fi*, will be present if a show is in class *C+* or *C-*, are then computed in accordance with Fig. 5b. These calculations can be performed once a day during times that the TV is not being viewed and stored in the viewer profile.

Recommendations for upcoming shows can be computed by estimating the posterior probabilities, i.e. the probability that a show is in class *C+* and *C-* given its features. Let x be a binary vector *(x1, x2, ...,xi, ..., xn)* where i indexes over the features in the viewer profile and where $xi = 1$ if feature *fi* is present in the show being considered for recommendation and 0 if not. For the exclusive features, like day, time, and station, where every show must have one and only one feature, the index i will be taken to indicate the value present in the show being considered provided that this value is also present in the profile. Otherwise, novel exclusive features will not enter into the calculations. For non-exclusive features, the index i will range over all values present in the profile; non-exclusive features novel to the considered show will not contribute to the calculations. The posterior probabilities are estimated in accordance with Fig. 5c

With these estimates in hand, a show will generally be recommended if $P(C+|x) > P(C-|x)$ and the "strength" of the recommendation will be proportional to $P(C+|x) - P(C-|x)$. One potential problem with this scheme is that some conditional probabilities are likely to be zero. Any zero in a chain multiplication will reduce the result to zero so some means for eliminating zeros is needed. The Billsus and Pazzani article referenced above presents a couple of schemes, including simply inserting a small constant for any zeros that occur.

One method for dealing with zeroes in the conditional probability multiplication chain would be as follows. One can choose a heuristic of 1000. If the number of shows in the viewing history is less than 1000, then the value of 1/1000 can be substituted for zero. If the number of shows in the viewing history is greater than 1000, the correction can be

$$\frac{k_i + \quad +1}{k + \quad +2}$$

Where

k_i+ is the number of watched shows having feature I

k+ is the total number of watched shows.

This is what is called the Laplace correction in the Billsus and Pazzani article. Alternative schemes may be devised by those of ordinary skill in the art.

Classical Bayesian theory would require the use of all accumulated elements of the list of Fig. 4 in making a recommendation. Nevertheless, in some instances it may be useful to use a noise cutoff, eliminating features from consideration if insufficient data about them appears in the list. For instance if a particularly feature did not appear in more than some given percentage of shows considered, whether in negative or in positive count, it might be ignored in determining which recommendation to make. Experimentally it was found that a cutoff of 5% was far too large.

Rather than use a percentage, one embodiment of the noise cutoff would use the viewer profile itself to determine the cutoff. This embodiment would first take a subset, or sub-list, of the viewer profile relating to particular feature types. For instance, a sub-list might advantageously comprise all of the elements of the viewer profile relating to the feature types: time of day and day of the week. Alternatively, in another example, the sub-list might advantageously comprise all of the elements of the viewer profile relating to channel number. Generally the feature type or types chosen should be independent feature types, in other words feature types which do not require another feature type to be meaningful.

The sub-list is then sorted by negative count, i.e. by number of shows having a particular feature value and not being watched. The highest negative count in this sorted list can be viewed as the noise level. In other words, since, in the preferred embodiment, the "not watched" shows are chosen at random from the week's program schedule, any not watched time slot can be considered to be noise.

Thus any feature having both a positive and a negative count at or below the noise level need not be considered in the Bayesian calculation in making a recommendation.

This example of noise level thresholding has used a particular feature, e.g. day/time as one for determining noise cutoff. In general, any feature that is uniformly randomly sampled by the negative example sampling procedure may be chosen by those of ordinary skill in the art for the calculation of the noise threshold.

The calculations of Figs. 5a-c are advantageous in that they require fairly low computing power to complete and are therefore readily adaptable to modest hardware such as would be found in a set top box.

"Surprise Me" Feature

Recommendations according to the above-described scheme will be programs having a preponderance of features that are present in shows that have been watched. The viewer profiles accumulated will not yield any meaningful recommendations with respect to shows having few features in common with watched shows. Accordingly, optionally, the recommender may occasionally recommend shows at random, in a "surprise me" feature, if such programs have relatively few features in common with watched shows.

Using the user profile in other domains

Once a user profile is developed, the recommendation techniques of the invention might be used to recommend other types of items such as movies, books, audio recordings, or even promotional materials such as tee shirts or posters.

Non-independence of Features

The classical assumption in the domain of Bayesian classifier theory is that all features are independent. Therefore, if a features is, say, often present in positive shows, but is missing from a show being considered for recommendation, the fact should count against the show. However, this may yield undesirable results for the current application.

For example, let us assume that there are five day/time slots indicated in the user profile as being most watched. Let us assume further that a particular show being evaluated falls within one of those five slots. The calculation of Fig. 5c would then give rise to an increase in probability for the day/time slot that matches and a decrease for the four day/time slots that do not match. Intuitively, it appears that the latter decrease is not reasonably related to an accurate determination of probability for the show in question. The different values of day/time are not independent -- as every show has one and only one value, so the values a show does not have should not count against it.

To remedy this deficiency in the classical Bayesian approach, it is proposed to designate features into two types: set 1 and set 2. If a feature is designated set 1, the Bayesian calculation will ignore any non-matching values of the feature. If the feature is designated set 2, then the normal Bayesian calculation, per Fig. 5c, will be done.

Normally in a television application set 1 would include day/time; station; and title. Some features which have values only for a few shows, e.g. critic ratings, should also be set 1, because too many shows would be non-matching merely because critics tend to rate only a tiny percentage of shows.

Set 2, for television shows, would normally include all features that can have several values per show, such as actor.

From reading the present disclosure, other modifications will be apparent to persons skilled in the art. Such modifications may involve other features which are already known in the design, manufacture and use of television interfaces and which may be used instead of or in addition to features already described herein. Although claims have been formulated in this application to particular combinations of features, it should be understood that the scope of the disclosure of the present application also includes any novel feature or novel combination of features disclosed herein either explicitly or implicitly or any generalization thereof, whether or not it mitigates any or all of the same technical problems as does the present invention. The applicants hereby give notice that new claims may be formulated to such features during the prosecution of the present application or any further application derived therefrom.

The word "comprising", "comprise", or "comprises" as used herein should not be viewed as excluding additional elements. The singular article "a" or "an" as used herein should not be viewed as excluding a plurality of elements.

CLAIMS:

1. A data processing device comprising:
- at least one input (35) receiving data including
  - viewer profile data; and
  - data regarding a television program; and
- a processor (2), the processor being adapted to perform the following
  - calculating a probability that the television program is a desired one; and
  - supplying a recommendation regarding the television program based on the probability.

2. The data processing device of claim 1 wherein the input (5) is coupled with a medium (110) readable by the data processing device.

3. The data processing device of claim 2, wherein the medium (110) embodies the viewer profile (103).

4. The data processing device of claim 3, wherein
the medium (110) is local to the data processing device and
the viewer profile (103) is arranged so as to be incrementally updatable.

5. The data processing device of claim 3, wherein the processor maintains the viewer profile (110, Fig. 4) in accordance with a data structure comprising:
a list of feature values; and
for each element of the list, a respective number of times programs having that feature value were watched.

6. The data processing device of claim 5, wherein the data structure further comprises, for each element of the list, a respective number of times programs having that feature value were not watched.

PNSDOCID: <WO 0158145A1 >

7. The data processing device of claim 6, wherein the processor is further arranged to perform the following, each time a user watches a new program,

first adding, to the list, feature values or counts of such feature values, associated with that new program;

selecting at least one companion program to the new program, the companion program being selected at random from a program schedule, which companion program has not been watched; and

second adding, to the list, feature values of the companion program, or counts of such feature values.

8. The data processing device of claim 5, wherein the processor is further arranged to perform the following, each time a user watches a new program: first adding, to the list, feature values or counts of such feature values, associated with that new program;

9. The data processing device of claim 2, wherein the medium (110) embodies the data (105) regarding the television program.

10. The data processing device of claim 1, wherein the input (5) is a network connection.

11. The data processing device of claim 1, wherein calculating comprises using a Bayesian classifier.

12. The data processing device of claim 11, wherein the processor is further adapted to subject the viewer profile (103) to a noise threshold calculation prior to using the Bayesian classifier.

13. The data processing device of claim 12, wherein
the viewer profile data (103) comprises
a list of feature values;

a respective negative count for each element of the list, the negative count indicating a number of times programs having that feature value have not been watched;

BNSDOCID: <WO 0158145A2 1 >

a respective positive count for each element of the list, the positive count indicating a number of times programs having that feature value have been watched;

the noise threshold calculation comprises

selecting a sub-list comprising at least feature values having at least one specific type of feature;

choosing the highest negative count in the sub-list as the noise threshold;

the recommendation comprises a program selected from a group having at least one feature value having a positive or negative count in the viewer profile, which count exceeds the noise threshold.

14. The data processing device of claim 12, wherein subjecting the viewer profile to the noise threshold further comprises using observations gathered by a known random process to estimate a reasonable noise threshold.

15. The data processing device of claim 13, wherein the specific type comprises a day and time of day feature type.

16. The data processing device of claim 13, wherein the specific type comprises a station identification feature type.

17. The data processing device of claim 1, wherein the viewer profile data (103) comprises a plurality of respective counts of programs watched, each respective count indicating how many programs watched had a respective feature.

18. The data processing device of claim 17, wherein calculating comprises calculating a probability that the television program is in a particular class.

19. The data processing device of claim 18, wherein the class is one of

programs the viewer is interested in, and

programs the viewer is not interested in.

20. The data processing device of claim 1, wherein calculating the probability comprises:

computing a prior possibility, of whether a program is desired or not;

computing a conditional probability of whether a feature fi will be present if a show is desired or not; and

computing a posterior probability of whether program is desired or not, based on the conditional probability and the prior probability.

21. The data processing device of claim 1, wherein it is assumed that programs watched are programs that the viewer is interested in.

22. The data processing device of claim 1, wherein

the viewer profile (103) comprises a list of features types and values for such feature types;

the feature types are selected from at least two sets, including

a first set of feature types whose values are deemed non-independent; and

a second set of feature types whose values are deemed independent; and

calculating a probability comprises

applying a Bayesian classifier calculation corresponding to feature types from the second set; and

applying a modified Bayesian classifier calculation corresponding to feature types from the first set.

23. The data processing device of claim 24, wherein

with respect to features of the first set, the modified Bayesian classifier calculation considers only feature values that match with a show being classified.

24. At least one medium (110) readable by a data processing device (2) and embodying software arranged to perform the following operations:

- calculating a probability that a television program is a desired one, based on a viewer profile (103) and data regarding the television program; and
- supplying a recommendation regarding the television program based on the probability.

25. The at least one medium (110) of claim 24, wherein the at least one medium (110) further embodies the viewer profile (103) as a data structure (Fig. 4) comprising:

a list of feature values; and

for each element of the list, a respective number of times programs having that feature value were watched.

26. The at least one medium (110) of claim 25, wherein the data structure (Fig. 4) further comprises, for each element of the list, a respective number of times programs having that feature value were not watched.

27. The at least one medium (110) of claim 26, wherein the software is further arranged to perform the following, each time a user watches a new program,

first adding, to the list, feature values or counts of such feature values, associated with that new program;

selecting at least one companion program to the new program, the companion program being selected at random from a program schedule, which companion program has not been watched; and

second adding, to the list, feature values of the companion program, or counts of such feature values.

28. The at least one medium of claim 24, wherein calculating comprises using a Bayesian classifier.

29. The at least one medium (110) of claim 28, wherein the software is further adapted to subject the viewer profile (103) to a noise threshold calculation prior to using the Bayesian classifier.

30. The at least one (110) medium of claim 29, wherein

the viewer profile data (103) comprises

a list of feature values;

a respective negative count for each element of the list, the negative count indicating a number of times programs having that feature value have not been watched;

a respective positive count for each element of the list, the positive count indicating a number of times programs having that feature value have been watched;

the noise threshold calculation comprises

selecting a sub-list comprising at least feature values having at least one specific type of feature;

choosing the highest negative count in the sub-list as the noise threshold;

the recommendation comprises a program selected from a group having at least one feature value having a positive or negative count in the viewer profile exceeding the noise threshold.

31. The at least one medium (110) of claim 24, wherein calculating the probability comprises:

computing a prior possibility, of whether a program is desired or not;

computing a conditional probability of whether a feature fi will be present if a show is desired; and

computing a posterior probability of whether program is desired or not, based on the conditional probability and the prior probability.

32. The at least one medium (110) of claim 24, wherein

the viewer profile (103) comprises a list of features types and values for such feature types;

the feature types are selected from at least two sets, including

a first set of feature types whose values are deemed non-independent; and

a second set of feature types whose values are deemed independent; and

calculating a probability comprises

applying a Bayesian classifier calculation corresponding to feature types from the second set; and

applying a modified Bayesian classifier calculation corresponding to feature types from the first set.

33. A computer method comprising performing the following operations in a data processing device:

Receiving a set of data;

Filtering the data in accordance with a noise criterion;

Drawing a conclusion from the filtered data based on a Bayesian classifier calculation; and

Presenting the conclusion to a user.

34. The method of claim 33 wherein the noise criterion is based on a frequency of instances of a particular types of data within set, which types are believed to likely represent noise.

35. A data processing method comprising performing the following operations in a data processing device:

- first receiving data reflecting physical observations, which data includes a list of feature values and observations about feature values, some of which feature values are independent and some of which are not;
- second receiving data about an item to be classified, the data about the item to be classified including feature values;
- maintaining a division of the data reflecting physical observations into at least two sets, including
  - a first set including those feature values which are deemed not independent; and
  - a second set including those feature values which are deemed independent;
- performing a probabilistic calculation on the data reflecting physical observations and the data regarding an item to be classified including:
  - applying a Bayesian classifier calculation with respect to feature values relating to the second set; and
  - applying a modified Bayesian classifier calculation with respect to feature values relating to the first set
  - presenting a conclusion regarding the item to be classified to a user based on the probabilistic calculation.

BNSDOCID: <WO 0158145A2 I >

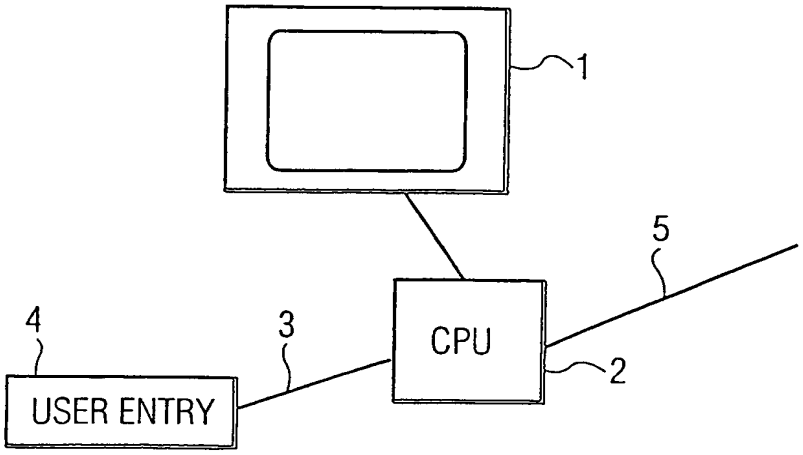


FIG. 0

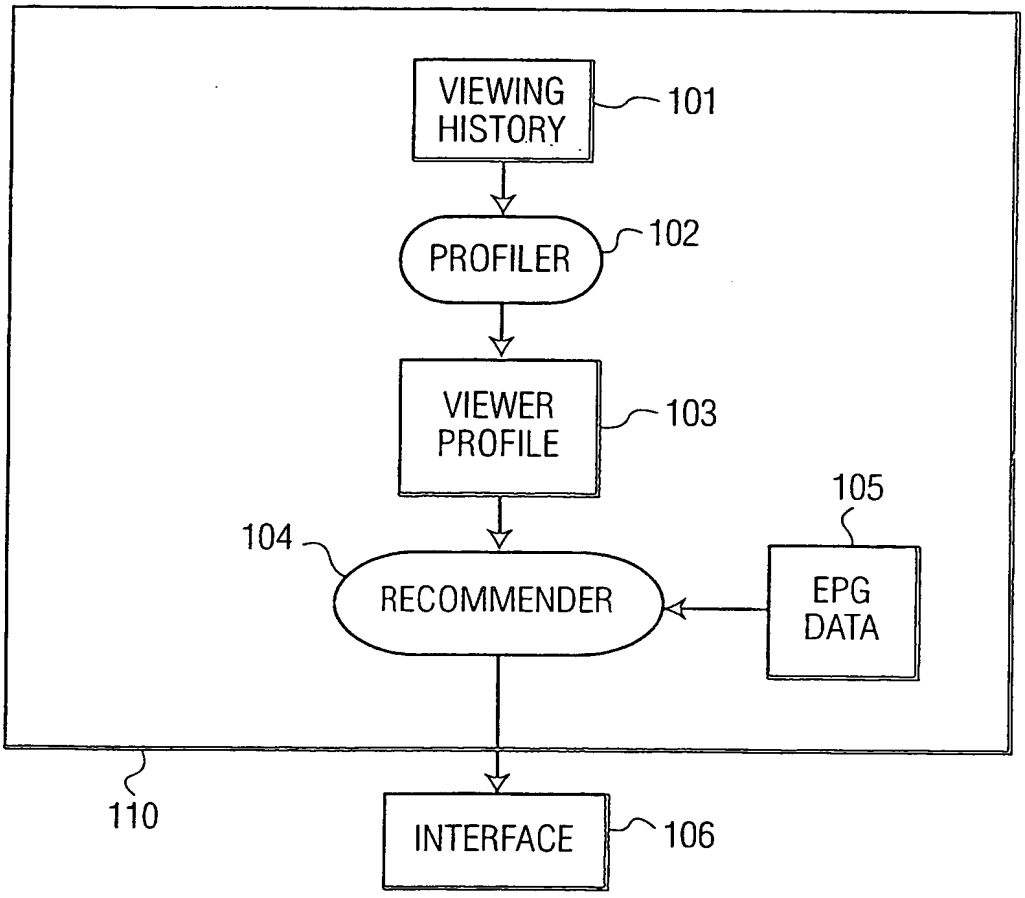


FIG. 1

BNSDOCID: <WO 0158145A2 >

```
LET S BE ALL SHOWS NOT WATCHED IN THE 7 DAYS PRIOR TO
      AND INCLUDING THE CURRENT DAY.
LET N BE 1.

FOR EACH TV SHOW WATCHED
      ENTER THE WATCHED SHOW IN THE VIEWING HISTORY
      AS A POSITIVE EXAMPLE.
      SELECT A SUBSET S OF SHOWS NOT WATCHED
      SELECT AT RANDOM N SHOWS FROM SET S AND ENTER
      THEM IN THE VIEWING HISTORY AS NEGATIVE EXAMPLES.
      IF AN EXPLICIT VIEWER PROFILE IS AVAILABLE, THEN THE
      RANDOM SELECTION CAN BE BIASED AWAY FROM SHOWS
      "LIKED" AND TOWARDS SHOWS "NOT LIKED."
```

FIG. 2

FIELD	DESCRIPTION
$DATE	YYYYMMDD
$AIR_TIME	HHMM (E.G. SOME VALUE IN THE RANGE 0000-2359)
$STATION_SIGN	4 CHARACTERS (E.G. WABC)
$TITLE	120 CHARACTERS (E.G. "ANTIQUES ROADSHOW")
$DESC	120 CHARACTERS (E.G. "SKULLY VISITS ALIEN SPACECRAFT")
$GENRE	20 CHARACTERS (E.G."SCIENCE FICTION")
$ACTORS	120 CHARACTERS (E.G. "JOHN DOE, JANE DOE")
$DIRECTORS	120 CHARACTERS (E.G. "JOHN HITCHCOCK")
$HOSTS	120 CHARACTERS (E.G. "JOHN HOST")
$PRODUCERS	120 CHARACTERS (E.G. "JANE RICH, JOHN MONEYBAGS")
$WRITERS	120 CHARACTERS (E.G. "JOHN POET")

FIG. 3

RNSDOCID: <WO 01581454A2 I >

TOTAL|PROGRAMS	55	55
DAYTIME|MON2100	5	0
DAYTIME|MON2200	6	1
DAYTIME|TUE2200	4	1
DAYTIME|WED2000	4	0
DAYTIME|WED2200	6	0
STATION|WABC	10	1
STATION|WNBC	30	0
STATION|WNYW	13	0
TITLE|20/20	5	0
TITLE|DATELINE NBC	11	0
TITLE|MLB PLAYOFFS	10	0
TITLE|PAID PROGRAMMING	0	5
GENRE|ANIMATED	0	4
GENRE|BASEBALL	13	1
GENRE|COMEDY	4	8
GENRE|FOOTBALL	4	0
GENRE|MAGAZINE	18	0
GENRE|NEWS	22	4
GENRE|REALITY	4	1
GENRE|SITUATION	3	5
GENRE|SPORTS EVENT	15	2
GENRE|TALK	18	3
DESC|ALLY	6	0
DESC|IF	6	0
DESC|MCBEAL	4	0
DESC|ALTERNATE	5	0
DESC|GAME	4	0
DESC|LINEUP	5	0

DESC|LOCAL	4	0
DESC|NECESSARY	6	0
DESC|PRIMETIME	5	0
DESC|PROGRAMMING	4	0
TV RATING|TV14	6	3
TV RATING|TVG	0	4
SEX RATING|N	51	52
LANGUAGE RATING|N	51	50
ADVISORY|ADULT SITUATIONS	0	8
ADVISORY|LANGUAGE	0	5
MPAA RATING|R	0	4
STAR RATING|**+	0	5
PROGRAM LANGUAGE|ENGLISH	51	52
COUNTRY OF ORIGIN|UNITED STATES	0	7
EPISODE TITLE|GAME	7	0
EPISODE TITLE|LEAGUE	4	0
EPISODE TITLE|NEW	5	0
EPISODE TITLE|SERIES	9	0

FIG. 4

BNSDOCID: <WO 0158145A2 I >

$$T = k(C+) + k(C-)$$
$$P(C+) = k(C+)/T$$
$$P(C-) = k(C-)/T$$

FIG. 5A

$$P(fi|C+) = k(fi|C+)/k(C+)$$
$$P(fi|C-) = k(fi|C-)/k(C-)$$

FIG. 5B

$$P(C+|x) = P(x|C+)P(C+)/P(x)$$
$$P(C-|x) = P(x|C-)P(C-)/P(x)$$

WHERE

$$P(x) = P(x|C+)P(C+) + P(x|C-)P(C-)$$

$$P(x|C+) = \prod_{i=1}^{n} P(fi|C+)^{xi} \ (1-P(fi|C+))^{1-xi}$$

n = number of features in profile
fi = the i^{th} feature in the profile is a bit string of length n, where the i^{th} bit indicates the
$x = \{0,1\}^n$ presence (1) or absence (0) of feature fi in the program

FIG. 5C

THIS PAGE BLANK (USPTO)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 August 2001 (09.08.2001)

PCT

(10) International Publication Number
WO 01/58145 A3

(51) International Patent Classification[7]: H04N 7/173, G06F 15/18, H04N 5/445

(21) International Application Number: PCT/EP01/00816

(22) International Filing Date: 25 January 2001 (25.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/498,271 4 February 2000 (04.02.2000) US

(71) Applicant: KONINKLIJKE PHILIPS ELECTRONICS N.V. [NL/NL]; Groenewoudseweg 1, NL-5621 BA Eindhoven (NL).

(72) Inventor: SCHAFFER, James, D.; Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(74) Agent: GROENENDAAL, Antonius, W., M.; Internationaal Octrooibureau B.V., Prof. Holstlaan 6, NL-5656 AA Eindhoven (NL).

(81) Designated States *(national)*: JP, KR.

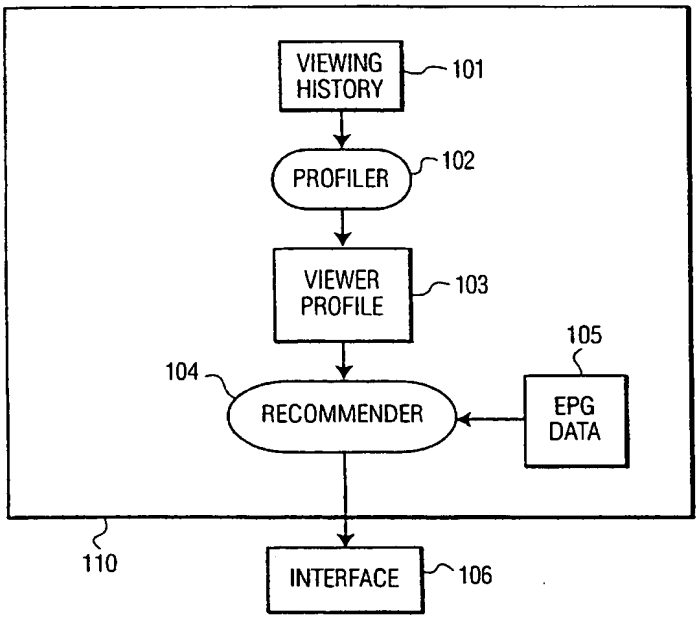
(84) Designated States *(regional)*: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR).

Published:
— *with international search report*

(88) Date of publication of the international search report:
10 January 2002

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: ADAPTIVE TV PROGRAM RECOMMENDER



(57) Abstract: A system for recommending television programs makes use of probabilistic calculations and a viewer profile to create a recommendation. The probabilistic calculations preferably are in the form of Bayesian classifier theory. Modifications to classical Bayesian classifier theory are proposed.

WO 01/58145 A3

INTERNATIONAL SEARCH REPORT

Inte· *ional Application No
PCI/EP 01/00816

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 H04N7/173 G06F15/18 H04N5/445

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 H04N G06F G10L

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, COMPENDEX, INSPEC, PAJ, IBM-TDB, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X Y	US 5 987 415 A (BREESE JOHN S ET AL) 16 November 1999 (1999-11-16) the whole document ---	33 1-32,34, 35
X Y	US 5 867 799 A (LANG ANDREW K ET AL) 2 February 1999 (1999-02-02) abstract; figures 1-6 column 6, line 66 -column 21, line 67 ---	33 1-32,34, 35
X Y	US 5 758 257 A (HERZ FREDERICK ET AL) 26 May 1998 (1998-05-26) abstract; figures 1-3,11 column 25, line 49 -column 30, line 40 --- -/--	1 2-32,34, 35

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
16 July 2001	31/07/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Brandenburg, J

Form PCT/ISA/210 (second sheet) (July 1992)

BNSDOCID: <WO 0158145A3 I >

INTERNATIONAL SEARCH REPORT

Inter ·ional Application No
PC1/EP 01/00816

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category ¨	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 410 344 A (BARKER ALVA C ET AL) 25 April 1995 (1995-04-25)	1
Y	the whole document ----	2-32,34, 35

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

BNSDOCID: <WO 01581454A1 >

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inter ional Application No
PCI/EP 01/00816

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5987415 A	16-11-1999	US 6185534 B	06-02-2001
		US 6212502 B	03-04-2001
US 5867799 A	02-02-1999	US 5983214 A	09-11-1999
		US 6029161 A	22-02-2000
US 5758257 A	26-05-1998	AU 703247 B	25-03-1999
		AU 4410396 A	19-06-1996
		CA 2207868 A	06-06-1996
		EP 0796538 A	24-09-1997
		US 6020883 A	01-02-2000
		WO 9617467 A	06-06-1996
		US 5734720 A	31-03-1998
		US 5754938 A	19-05-1998
		US 5754939 A	19-05-1998
		US 5835087 A	10-11-1998
		US 6088722 A	11-07-2000
		US 6029195 A	22-02-2000
US 5410344 A	25-04-1995	NONE	

BNSDOCID: <WO 0158145A3 I >